Since both $\rho(=\frac{m}{k})$ and $z(=\frac{1-\varepsilon}{\varepsilon}(\frac{E\rho(1-\alpha)}{(1-\beta)(E+\rho\alpha)})$ go to zero as K goes to infinity, in the limit the budget depends only on Π_3. When the monitoring probability in group 3 is taken to be 1, as it is here, the repeated play budget approaches $\frac{\alpha(1-\varepsilon)}{1+\alpha}$ in the limit. Therefore, the budget cannot be driven to zero by choosing an arbitrarily large K. And the limiting value of the ratio r_{sp}/r_{mp} is $(\frac{m}{1-\alpha})(\frac{1+\alpha}{\alpha})$. For example, for the first case in table 3, where $\varepsilon = 0.2$ and $\alpha = 0.05$,

$$\lim_{k\to\infty} \Pi_3 = \lim_{k\to\infty} r_{mp} = \frac{0.5(0.8)}{1.05} = 0.038$$

and $$\lim_{k\to\infty} (r_{sp}/r_{mp}) = \frac{0.585}{0.95}\cdot\frac{1.05}{0.05} = 12.93.$$

- Finally, higher fines imply a smaller budget advantage for multiple-play over single-play structures and the larger the errors, the smaller the advantage for constant fine size.

Concluding Comments

This paper makes a preliminary case for using game theory notions, especially the repeated game model, in the design of monitoring and enforcement systems. Most importantly, desired levels of compliance may be obtained with smaller budget allocations for monitoring when information on past compliance is used to define future monitoring probabilities. The relevance of these results for actual policy will depend on the resolution of several questions, some involving law and some extensions of the economics considered here.

The most important legal point is likely to be whether or not an arbitrary probability of escape from group 3, E, can in fact be applied to pollution sources (or any other monitored parties).

By far the most important economic question is how the results derived above are affected by the introduction of discounting at a positive interest rate. There is no question that discounting will reduce the effectiveness of the scheme, for the threat of a given stay in group 3 will be lower, everything else equal. The question is, by how much? Will the required monitoring budget, given a required upper limit on noncompliance, an error structure, and a fine size, be less than that implied in a single-play approach? Will it be enough less to make the relatively elaborate grouping mechanism worthwhile?

FOOTNOTES

1. It is an interesting and challenging feature of air and water pollution discharges that they are fugitive events in the sense that they must be measured as they occur or they cannot be measured at all. (There are statistical techniques for inferring air pollution discharges from ambient pollutant concentrations. (See Gordon 1980). That is, enforcing pollution control permit terms is more closely analogous to enforcing auto speed limits than, say, the tax laws. The situation for containerized hazardous wastes is different, of course.

2. In fact, most monitoring visits are the result of the application of rules of thumb such as: Visit large sources once every N months. Such visits usually also involve prior announcement for convenience and to avoid legal battles over rights of entry. Thus, the deterence effect of the visits is even less than that implied by the rates.

3. One visit is assumed to be the maximum required to guarantee compliance in a period (subject to the qualifications implied by errors of inference). That is, if a monitoring visit occurs and a violation is detected, that violation is assumed to be corrected for the remainder of the period. It is further assumed that the full costs of compliance in the period, C, will be incurred if a source in violation is caught and forced into compliance. A false violation notice (false positive) implies that a fine is levied on a source already spending C to comply.

4. A natural question is, why not announce that m applies to all sources, but actually only apply r? As a short-term strategy of deceit this may seem appealing. But it is conceptually weak reed on which to build a policy, for assuming that deceit can successfully be practiced indefinitely seems to imply zero information gathering activity and inferential ability on the part of the sources. Similarly, relying on an announced monitoring probability, m, applying to an unknown fraction of sources $\frac{N_r}{N}$ gives an effective monitoring probability for a source of:

$$r = \frac{N_r . m}{N} = \frac{\frac{N.r.m.}{m}}{N} .$$

It seems wise to assume that each source will infer the actual probability it faces, and therefore no sources will have the incentive to comply.

5. If $\alpha = 0$, the measurements are effectively perfect for these purposes and $r = (1 - \varepsilon)m$.

6. Proving that it is optimal for a source to comply in group 2 while violating in group 1 involves defining the following: $E(V_2)$ is the expected value of the strategy violating when in group 2; $E(C_2)$ is the expected value of the strategy of complying in group 2, and $E(V_1, C_2)$ in the expected value of violating in group 1 and complying in group 2. Observe the following:

$$E(V_1, C_2) = (1 - z)0 + zE(C_2)$$

$$E(C_2) = C + \rho(E(V_1, C_2)) + (1-\rho)(E(C_2))$$

$$E(V_2) = 0 + \rho \sum_{t=1}^{\infty} C_t + (1 - \rho)E(V_2)$$

so that:

$$E(V_2) = \sum_{t=1}^{\infty} C_t \to \infty$$

$$E(C_2) = C + \rho(zE(C_2)) + (1 - \rho)E(C_2)$$

$$E(C_2)[1 - \rho z - (1 - \rho)] = C$$

$$E(C_2) = \frac{C}{\rho - \rho z}$$

And so long as $z < 1$, this number is positive but finite, so that $E(C_2) < E(V_2)$. $z < 1$ is guaranteed by the rule proposed in the text for defining z in terms of ε and ρ.

REFERENCES

Becker, Gary S. 1968. "Crime and Punishment: An Economic Analysis," Journal of Political Economy, 76 (March/April), pp. 169-217.

Casey, Don, Peter M. Nemetz, and Dean H. Uyemo. 1983. "Sampling Frequency for Water Quality Monitoring: Measures of Effectiveness", Water Resources Research. vol. 19. no. 5. (October) pp. 1107-1110.

Fortuna, Ricard C. and David J. Lennett. 1987. Hazardous Waste Regulation: The New Era., (McGraw Hill, NY).

Gordon, Glen E. 1980. "Receptor Models" Environmental Science and Technology. vol. 14, no. 1, pp. 792-800.

Greenberg, Joseph. 1984. "Avoiding Tax Avoidance: A (Repeated) Game-Theoretic Approach," Journal of Economic Theory, vol. 32, no. 1, (Feb.), pp. 1-13.

Landsberger, Michael, and Isaac Meilijson. 1982. "Incentive Generating State Dependent Penalty System," Journal of Public Economics, vol. 19, pp. 333-352.

McKean, Roland. 1980. "Enforcement Costs in Environmental and Safety Regulation," Policy Analysis, vol. 6, no. 3, pp. 269-289.

Russell, Clifford S. 1982. Pollution Monitoring Survey Summary Report, (Washington: Resources for the Future).

Russell, Clifford S. 1987. forthcoming, "Monitoring and Enforcement," in Paul R. Portney, (ed.) Environmental Regulation In The U.S.: Public Policies and Their Consequences. (Washington, DC: Resources for the Future).

Russell, Clifford S., Winston Harrington and William J. Vaughan. 1986. Monitoring and Enforcing Pollution Control Laws. (Washington, DC: Resources for the Future).

U.S. GAO, 1979. Improvements Needed in Controlling Major Air Pollution. (Washington, DC: U.S. GAO) January, CED-78-165.

U.S. GAO, 1980. Costly Wastewater Treatment Plants Fail to Perform as Expected. (Washington, DC: U.S. GAO) November CED-81-9.

U.S. GAO, 1983. Wastewater Dischargers Are Not Complying With EPA Pollution Control Permits. (Washington, DC: U.S. GAO) December GAO/RCED-84-53.

U.S. GAO, 1985. Air Pollution: Environmental Protection Agency's Inspections of Stationary Sources. (Washington, DC: U.S. GAO) Oct. GAO/RCED-86-1BR.

U.S. Nuclear Regulatory Commission, Office of Inspection and Enforcement. 1983. Inspection and Enforcement Manual, chapter 0300 (January) Washington, DC.

Wasserman, Cheryl. 1984. "Improving the Efficiency and Effectiveness of Compliance Monitoring and Enforcement of Environmental Policies: United States: A National Review," prepared for the Organization for Economic Co-operation and Development, Environmental Directorate, (October). Washington, DC.

The Behavior of the Firm When Facing Uncertain Enforcement of Ex Ante Environmental Regulation and Tort Liability Standards

by

Gary V. Johnson
Institute for Environmental Studies
and
Department of Agricultural Economics

and

Thomas S. Ulen
Department of Economics
and
Institute of Government and Public Affairs
University of Illinois at Urbana-Champaign

Prepared for the AERE Workshop on
Environmental Enforcement and Monitoring
Newark, Delaware
August, 1987

Acknowledgements: We wish to thank Juha Alho, Institute for Environmental Studies, University of Illinois at Urbana-Champaign for his comments on an earlier draft and the Research Board of the University of Illinois for their generous financial assistance.

The economic policy analysis of environmental issues has generally focused on the relative efficiency in minimizing environmental harms of the unregulated market and of administrative agency (or what we call ex ante) regulation. Once the scholarly analysis has indicated the circumstances that dictate governmental intervention in the market, it has then focused on the relative efficiency of different forms of this ex ante regulation. For example, the standard reliance on a system of fixed fines for violation of environmental regulations has been challenged as inefficient by comparison to a system of sliding effluent fees or to a requirement that potential polluters install the best available (in the sense of least-polluting) technology for production of their output.

Recently, economists and lawyers familiar with economics have explored the efficiency of exposure to tort liability (what we will call ex post regulation) as part of a policy for minimizing external costs such as environmental harms.[1] The premise is that potential polluters can be constrained to produce an efficient amount of pollution by making them liable in a private cause of action for compensating those whom they have harmed. Until very recently this method of regulating environmental harms has been a rare exception.[2]

This article attempts to bring together these two strands of the environmental and the law and economics literatures in order to investigate the relative efficiency of ex ante and ex post regulation. The question

---

[1]See Robert D. Cooter and Thomas S. Ulen, *Law and Economics* (Scott, Foresman and Co.: Glenview, IL, 1987), pp. 326-371.

[2]For exceptions, see Boomer v. Atlantic Cement Co., Inc., 26 N.Y.2d 219, 309 N.Y.S.2d 312, 257 N.E.2d 870 (Court of Appeals of New York, 1970), and Spur Industries, Inc. v. Del E. Webb Development Co., 494 P.2d 701 (Ariz. 1972). The economics of the cases are discussed in Cooter and Ulen, op. cit., 171-185.

naturally raised by this conjunction of the regulatory literature is whether *ex ante* and ex post regulation are complements or substitutes. Educated economic intuition would suggest that they are substitutes; either alone would seem to be able to accomplish the goal of creating efficient incentives but to use them jointly would seem to be wasteful duplication. It turns out that this intuition is correct only when there is no uncertainty in the enforcement of either the administrative agency standard or the tort liability standard. In that unlikely case, the two methods of regulation are, indeed, substitutes, and the social costs of regulating the externality are minimized by relying exclusively on whichever of the two forms of regulation has the lower administration and enforcement costs. The far more likely case is that there is uncertainty in the enforcement of either or both forms of regulation (Russel *et al.).* Where that is true, then ex ante and ex post regulation become complementary regulatory tools. We have recently shown that the joint use of these forms of regulation creates efficient incentives for potential externality-generators when there is certainty regarding the enforcement of the ex ante regulatory standard but so much uncertainty regarding the enforcement of the ex post liability rule that firms take less precaution to prevent environmental harms than is socially optimal (Johnson *et al.).* In this article we extend those results by considering the case where there is uncertainty in the enforcement of both forms of regulation.

The policy conclusions that result from analyzing the effect of uncertain enforcement in both forms of regulation can be readily stated. The existence of uncertain enforcement of both an ex ante administrative agency regulation and an *ex post* liability rule for a given environmental externality strengthens the case for the simultaneous use of these two forms of regulation but slightly alters the relationship between the two. When only the enforcement of the *ex post* standard is uncertain, the (certain) *ex ante* standard should, surprisingly, be set below the socially optimal level of care. When there is uncertainty in the enforcement of both standards, it is still true that the *ex ante* standard should be set below the *ex post* standard, which, by assumption, is set equal to the socially optimal level of care. But now the distance between the two standards depends on the relative degree of uncertainty in the enforcement of the two standards in an intuitively plausible way that we explain below.

The remainder of the paper is organized in the following way. In Section II we discuss the economic analysis of tort liability and apply that analysis to the policy issues created by the generation and disposal of hazardous wastes. Section III develops the formal model of the firm's behavior when faced with both *ex ante* and *ex post* regulation of potential environmental harm with special attention to the impact of the uncertain enforcement of both of those standards on the potential polluter's behavior. The paper concludes with remarks on two matters. First, we suggest the implications of the model for public policymaking regarding environmental harms. Second, we speculate on the next steps in the theory of integrating *ex ante* and *ex post* regulation and on the empirical work that our formal model suggests.

II. The Economics of Tort Liability

The economic analysis of tort liability standards is sufficiently new and arcane that a brief introduction is in order. Recall that the simple economic premise is that the potential injurer and victim will be induced to take optimal precaution against harm if they are liable for compensating those whom they harm. This potential liability becomes a part of the decisionmaker's anticipated costs, which he then attempts to minimize by taking the optimal amount of precaution. The common law recognizes two different tort liability standards—negligence and strict liability. But the law has not provided a convincing explanation of the circumstances in which negligence is superior to strict liability and *vice versa.* The economic analysis of tort liability *has* provided such an explanation. Let us briefly summarize this explanation. It is an important one in designing public policy for environmental harms and in introducing the formal model of the next section.

The tort liability standard of negligence establishes a legal standard of care or precaution that the potential injurer and victim owe to one another.[3] If a party violates the legal standard by taking less care or precaution than he owes to the other party and that failure to take care proximately caused the harm, then he is negligent or at fault and is liable for the losses suffered by the victim; if he takes at least as much care as required by the legal standard or if his actions did not proximately cause harm, even though he was negligent, then he is not negligent or at fault and is not liable for the victim's losses. The negligence standard comes in several varieties—simple negligence, negligence with contributory negligence, and comparative negligence—that differ in the extent to which they take into account the victim's own precaution against the occurrence of a harm. The general conclusion of the literature on the efficiency aspects of the negligence rule is that under certain

[3]The following material draws on Cooter and Ulen, op. cit., 326-371.

conditions all forms of the negligence rule create equally efficient incentives for potential victims and injurers to take precaution and, therefore, minimize the social costs of external harms. The conditions under which this conclusion holds are that (1) the legal standard of care is set equal to the socially optimal level of care, (2) precaution is "bilateral," and (3) there is no imperfection or uncertainty in determining whether or not a party complied with the legal standard of care. By "bilateral precaution" we mean that the technology of precaution is such that both parties may reasonably take action to reduce the probability or severity of the external harm. An example would be automobile accidents: typically there is something that both drivers can do to reduce the probability or severity of an accident, even if in any given accident only one of them was in fact at fault. A point worth emphasizing here is that *all* forms of the negligence rule are equally efficient when there is no uncertainty in the enforcement of the legal standard of care.

When we introduce uncertain enforcement into the economic analysis of negligence, the conclusion changes. Suppose that enforcement is uncertain in the following sense: it is possible that (1) a party who complied with the legal duty of care is nonetheless held liable or (2) a party who failed to comply with the legal duty of care is not held liable. Under this uncertainty there may be important efficiency differences among the forms of the negligence rule (Cooter and Ulen, 1986; Craswell and Calfee, 1984 and 1986.). This highlights the importance of uncertainty in modeling behavior under ex post regulation.

The other great liability standard is strict liability. Under that standard, an injurer is liable for the victim's losses if he proximately caused the harm. The potential injurer is under an absolute duty not to cause harm; there is no legal standard of care or precaution that he may take to exonerate him from liability for the victim's losses. Moreover, there is usually no requirement that the potential victim himself take care; the entire burden is on the potential injurer. Strict liability creates efficient incentives for precaution by the potential injurer under three conditions: (1)when he is liable to the victim for "perfect" compensation, (2) when precaution is "unilateral," and (3) when there is no uncertainty in the enforcement of the strict liability standard. Perfect compensation is an award of money damages to the victim that makes him indifferent between the state of having suffered the harm but receiving the money damages and that of never having suffered the harm. Precaution is said to be unilateral when only the potential injurer may reasonably take action to reduce the probability or severity of harm. Uncertain enforcement can arise under strict liability in two ways. First, the determination of proximate cause may be subject to uncertainty: an injurer may be held strictly liable when in fact his actions did *not* proximately cause harm or he may be excused from liability when in fact his actions *did* proximately cause the injury. Second, the victim's compensation can be imperfectly measured. Potential injurers thus may be liable for widely-varying amounts of compensation: some victims whose loss was minimal may be vastly over-compensated while others whose loss was extraordinarily large may be under-compensated. Where there is uncertain enforcement in either or both of these senses, the efficiency of strict liability is lessened.

The uncertain enforceability of tort liability standards clearly lessens their efficiency. (We assume that the other conditions specifying whether strict liability or negligence is the appropriate liability rule are met; that is, the only source of inefficiency upon which we wish to focus is uncertain enforcement; for instance, strict liability is not being applied in circumstances of bilateral precaution, nor is negligence being applied in cases of unilateral precaution.) The question that this lessened efficiency raises is whether there is some way to improve the efficiency of an uncertainly-enforced *ex post* regulatory rule. There are several possibilities:

(1) depending on the direction of the inefficiency (i.e., whether it leads to too much or too little precaution), the rules for establishing fault or causation could be relaxed or tightened;

(2) assuming that there is so much uncertainty that potential injurers inevitably take too little precaution, courts could routinely award punitive damages in addition to compensatory damages) in those instances where an injurer is held liable;

(3) if the inefficiency is extraordinarily large, some alternative regulatory tool for minimizing the external harm might be substituted (e.g., the activity could simply be outlawed or victims could be compensated in an administered compensation or no-fault system);

(4) a complementary regulatory tool could be used (e.g., in addition to exposing potential wrongdoers to tort liability, they could also be exposed to siting requirements or quality controls backed by fines.

To focus the discussion let us apply the foregoing analysis to a concrete problem in environmental regulation: harms arising from exposure to hazardous wastes. Let us first briefly summarize, for future use, the federal and state statutory regulations dealing with that problem and then discuss how the economic analysis of tort liability might be applied to this issue.

Federal statutory regulation of hazardous wastes began with the Safe Drinking Water Act.of 1974.[4] Among other things that Act told the US. Environmental Protection Agency (EPA) to require the states to promulgate regulations dealing with the underground injection of hazardous wastes. Two years later Congress passed the Resource Conservation and Recovery Act of 1976 (RCRA).[5] That Act provided for tracking hazardous wastes from the time and point of generation until final disposal but failed to provide for any problems associated with improper disposal before 1976. In 1980 Congress attempted to correct this failure in the Comprehensive Environmental Response, Compensation, and Liability Act (CERCLA).[6] That Act empowered the Federal government through the EPA to impose liability for cleaning up hazardous waste disposal sites closed before the enactment of RCRA.[7] CERCLA has been revised but its basic structure remains intact.

State statutory regulation is predictably less coherent. Some states have "mini-superfunds,"[8] and many have additional, more direct controls. For example, Illinois and Massachusetts have banned landfilling, the most popular and economical method of disposing of hazardous wastes;

---

[4] 42 U.S.C. secs. 300f-300j (1982).

[5] 42 U.S.C. secs. 6901-6907, 6911-6916, 6921-6931, 6941-6954, 6961-6964, 6971-6979, 6981-6986 (1982).

[6] 42 U.S.C. secs 9601-9657 (1982). CERCLA created the "Superfund," a fund financed by Congressional appropriations and taxes on those who produce hazardous and toxic wastes. The EPA is empowered to use this fund to clean up dangerous waste sites and then to bring actions against various parties for recovery of the cleanup costs under sec. 107 of the Act.

[7] For an excellent summary of the law on hazardous wastes, see "Developments in the Law -- Toxic Waste Litigation," 99 *Harv. L. Rev.* 1459 (1986).

[8] For an example see the Minnesota Environmental Response and Liability Act (MERLA), Minn. Stat. Ann. §§115B.01-115B.24 (West 1985 supp.).

some states have extended the reporting requirements of RCRA to small firms (such as dry cleaners and gasoline stations) that are exempt from the federal regulations.

The common law tort liability treatment of harms arising from the generation and disposal of hazardous wastes is in its infancy. Only a handful of actions have been filed so that it is not clear how the common law will treat these matters. For that reason we may begin at the beginning and explore how the common law *should* deal with hazardous wastes by applying the economic analysis of tort liability sketched above.

Consider how the distinction between negligence and strict liability applies to the environmental harms that might arise from the generation or disposal of hazardous wastes. For those harms, precaution is certainly unilateral in the sense that disposers and generators are the only parties to whom society may reasonably look for actions that will reduce the probability or severity of harm from those sources. Thus, if this harm is to be regulated using an *ex post* liability rule, injurers (the disposers and perhaps the generators of hazardous wastes) should be held strictly liable to those they have harmed.

But what about the problems of uncertain enforcement under strict liability?[9] Recall that in making a claim for recovery under strict liability, the victim must show that the injurer proximately caused the harm (but not that in doing so the injurer violated a standard of due care owed to the plaintiff). Two problems are likely to arise in establishing

[9] The following discussion comes from Gary V. Johnson and Thomas S. Ulen, "Designing Public Policy Toward Hazardous Wastes: The Role of Administrative Regulations and Legal Liability Rules," *American Journal of Agricultural Economics* (Dec. 1986) pp. 1269-1270.

that the generation or disposal of hazardous waste has proximately caused a harm. First, the scientific evidence on causation between exposure to these substances and personal or property injury is at an early stage of development. Thus, it may well be the case that although by the standards of the academic community the causal connection may be reasonably well established, by the standards of proof required in a court the causal connection is not clear enough to permit recovery in amounts close to perfect compensation (Farber, Davis). Second, harms arising from the generation or disposal of hazardous wastes may not become manifest for long periods of time, sometimes more than a generation. If so, the evidence necessary to establish proximate cause for recovery under a strict liability theory may be so distant or so clouded that otherwise meritorious plaintiffs cannot recover. Where this is the case, then the generators and disposers of hazardous wastes do not receive the appropriate signal from the tort liability system about the appropriate level of precaution to take, and as a consequence they may take too little precaution.

To what extent can the strict liability standard be amended to take account of these special problems of recovering for harms arising from the generation and disposal of hazardous wastes? With regard to the problem of establishing proximate cause where scientific evidence is weak, there are several possibilities. Legislation might relax the plaintiff's burden of proof on the causation issue. This is the tack adopted in the 1983 Minnesota Environmental Response and Liability Act (MERLA) and proposed by some commentators for widespread adoption.

This relaxation of the plaintiff's burden of proof on the causation issue for harms arising from the generation or disposal of hazardous and toxic wastes may cure the first problem noted above, but it does so at a high cost. Tampering with the traditional causation standard is a radical step that requires extraordinary justification, a justification that no one has offered. Naturally the question arises: If the inadequacy of scientific knowledge to establish proximate cause in the case of harms arising from hazardous and toxic wastes justifies relaxing the plaintiff's burden of proof in those harms, why may we not also relax plaintiffs' burdens in all other instances of inadequate scientific knowledge? Thus far, there has been no good answer to that question. And that fact suggests that relaxation of the plaintiff's causation requirement is not yet an acceptable method of making strict liability for harms inflicted by exposure to hazardous and toxic wastes more efficient.

The second source of uncertain enforcement of strict liability for harms associated with hazardous wastes is the long time lag between exposure and manifestation of the harm. This lag complicates the injurers ability to demonstrate proximate cause and, therefore, makes his recovery, even if his case is meritorious, less likely. This problem has arisen in several well-known modern cases, e.g., in the diethylstilbestrol (DES) and asbestos cases. Some commentators have suggested that the tort liability system can be reformed to accommodate the peculiar evidentiary problems of time-delayed harms by allowing probable victims to recover from a probable injurer before any actual harm has become manifest. The proposal is that where the probability of any harms developing in the future is above some minimum threshold, the potential victim should be allowed to recover the expected damages discounted by the probability that the harm will arise (Cooter and Ulen, 1987).

It is argued that the benefit of allowing recovery under a theory of probabilistic causation for inflicting tortious risk is that the signal to take efficient precaution will be transmitted relatively quickly to potential injurers. Not allowing recovery in these circumstances, it is argued, will greatly reduce the number of cases that plaintiffs can win in the distant future and consequently will greatly dilute the signal to injurers to take efficient precaution.

But the costs of revamping tort law to allow for probabilistic causation and recovery for infliction of tortious risk are also high. One of the most fundamental precepts of tort law is that a harm must have occurred; simply creating a dangerous condition--what has been called, in a famous phrase, "negligence in the air"--is not a sufficient basis ford bring an action. There are good efficiency reasons for limiting recovery in tort to cases of actual harm. Moreover, there are almost insurmountable problems involved in specifying the threshold probability of harm that would trigger liability.

These observations suggest that the uncertain enforcement that is likely to arise under a strict liability standard for harms arising from exposure to hazardous wastes cannot be easily corrected with the ex post regulation system itself.

It is still an open question whether an uncertainly-enforced tort liability standard is best supplemented or replaced by uncertainly-enforced federal and state administrative agency regulation and like those we described above. We turn to that question in the next section with the help of a formal model of firm behavior under uncertainly-enforced ex ante and ex post regulation.

III. Modeling the Firm's Behavior Under Uncertain Enforcement of Regulation

To examine the effect on the firm's behavior of uncertainty regarding the enforcement of both ex ante environmental regulation and ex post liability, we will first look at a simple model of firm behavior and enrich the model by adding greater complexity. We begin our model development by reviewing a deterministic model of the firm behavior when faced with strict liability that is common to the law and economics literature (Cooter and Ulen, 1987). Strict liability is chosen because of the unilateral nature of precaution regarding many environmental harms. Uncertainty regarding the enforcement of a strict liability rule will then be introduced. We will examine the effects of assumptions concerning the distribution of that uncertainty and the nature of the cost function associated with precaution. Next, we will focus on *ex ante* regulation, beginning with a simple deterministic model of a regulatory standard. Uncertainty regarding the enforcement of the standard will then be introduced and its impacts explored. Finally, we model the firm's behavior under the joint use of *ex ante* regulation and *ex post* liability when enforcement of both forms of regulation is uncertain.

A. A Model of the Firm Facing Strict Liability

Let x be the level of the firm's precaution in preventing an environmental harm. The costs of precaution are given by the function $C(x)$, which is upward sloping, i.e., $C'(x) > 0$, and convex over the relevant region. The expected size of the harm that the firm can anticipate is given by the function $A(x)$, which is assumed to be convex and downward sloping over the relevant region, i.e., $A'(x) < 0$. [$A(x)$ embodies both the severity of the harm and the probability of its occurring.]

The expected social costs of the externality or harm are the sum of the precautionary costs, $C(x)$, and the expected costs of the harm, $A(x)$. Assume that this sum, $[C(x) + A(x)]$, is strictly convex. The socially optimal amount of precaution, x^*, is that which minimizes these social costs, i.e.,

$$\min_{x} SC(x) = [C(x) + A(x)], \tag{1}$$

where $SC(x)$ is the total social cost. At x^* the marginal expected cost of precaution equals the negative of the expected marginal cost of the harm, i.e.,

$$C'(x^*) = -A'(x^*). \tag{2}$$

It can be shown that in the absence of uncertainty, with perfect victim compensation and unilateral precaution, the firm subject to strict liability will choose the socially optimal amount of precaution, x^* (Cooter and Ulen, 1987).

As we noted informally in the previous section, if uncertainty exists in the determination of proximate cause or in the determination of perfectly compensatory damages for the victim, then the firm may choose a socially inefficient amount of precaution (Johnson *et al.).* To simplify the modeling of enforcement uncertainty regarding the firm's choice of the level of precaution, we assume for the moment that victim compensation is perfect. Given this assumption, the focus becomes the logically prior issue of causality. The firm perceives enforcement uncertainty regarding whether the court will find that its actions proximately caused the victim's harms. This could arise, as noted above, from a lack of the scientific knowledge necessary to establish proximate cause. Let q(x) be the firm's subjective probability distribution around what it perceives as its cost-minimizing level of precaution, $\hat{x}$.[10] Assume that q(x) is a continuous probability density with support $[0,\infty)$. The probability of the firm's actions being deemed the proximate cause of the harm and of the firm's therefore being held strictly liable to the victim is thus given by

$$R(x) = \int_{x}^{\infty} q(x)dx. \tag{3}$$

That is, R(x) is the probability that the firm will be held liable to pay (perfectly compensatory) damages, given a precaution level of amount x.

How will the firm respond to this form of uncertainty? The firm's objective function can be formulated as the minimization of its precautionary and expected liability costs:

$$\text{Total Costs} = TC(x) = d(x) + A(x)R(x). \tag{4}$$

Previously we assumed that [C(x) + A(x)] is strictly convex. Now we assume that [C(x) + A(x)R(x)] is strictly convex. The essence of the model is presented in Figure 1. The socially optimal level of precaution, x^*, is where the marginal precaution costs just equal the negative of the marginal expected harm. When certainty exists in the determination of proximate cause, the firm's optimal amount of precaution will coincide with the social optimum, x*. But where uncertainty exists in determining causality, then the probability that the firm will be held liable at any

---

[10] For the sake of simplicity, we will assume that $\hat{x}$ and x*, the social-cost-minimizing level of precaution, are equal. Later we will analyze the case where $\hat{x}$ and x* differ.

given level of precaution, say $\bar{x}$, is the area $R(\bar{x})$ under the distribution and to the right of $\bar{x}$.

It can be shown that if TC(x) is strictly convex, this function has a unique minimum (Johnson *et al.).* The first-order condition for the minimization is

$$C'(\tilde{x}) + A'(\tilde{x})R(\tilde{x}) - A(\tilde{x})q(\tilde{x}) = TC'(\tilde{x}) = 0. \tag{5}$$

Equation 5 is basic to the subsequent analysis and thus deserves some interpretation. The first term on the left-hand-side of the equation is the marginal cost of providing a unit of precaution. The second and third terms sum to the expected marginal liability costs of a unit of precaution and consist of two effects. The first effect, $[A'(\tilde{x})R(\tilde{x})]$, is the marginal expected cost of the harm times the probability of being held liable for the harm if the firm's level of precaution is equal to ii. This term, which we call the "injury effect," is negative because $A'(\tilde{x})$ is negative (more precaution reduces the severity and probability of harms) and $R(\tilde{x})$ is always positive. But there is also a savings from providing slightly higher precaution, a reduction in the probability of being held liable. This savings is captured in the term $[-A(\tilde{x})q(\tilde{x})]$, which we call the "liability effect." This term is negative because $-A(\tilde{x})$ is negative and, by assumption, $q(\tilde{x})$ is positive. It follows that the firm minimizes its expected costs by taking the precaution for which the liability marginal cost is equal to the effect minus the injury effect.

Whether a little more or less precaution is better at minimizing the firm's costs depends on the marginal cost of precaution and on the relative size of the liability and injury effects. We cannot be certain of this without making some additional assumptions about the components of TC(x).

For present purposes, the important assumptions concern the nature of the distribution q(x) and the behavior of the marginal cost function C'(x).

Turning first to assumptions concerning the nature of the distribution q(x), we can show (see Johnson *et al.)* the following two propositions.

Proposition 1. Assuming a mean-preserving spread, an increase (decrease) in the uncertainty surrounding the legal standard will result in firms' reducing (increasing) their level of precaution. If the uncertainty in determining causality is great (small) enough, firms will take too little (much) precaution ***vis-á-vis*** the social optimum.

This case is shown in Figure 2.[11] This figure demonstrates that even if the firm's expected value of precaution is x* (the social optimum) the presence of a sufficient degree of uncertainty will result in under- or over-precaution.

Proposition 2. If the firm's subjective distribution is biased so that its mean is below (above) the social optimum, then the firm also will take too little (much) precaution.

The last case is shown in Figure 3. Note that as the distribution shifts to the right (left) the variance of the distribution increases (decreases). This need not be the case. Examination of Proposition 1 suggests that if the shift in the distribution preserved variance, then Proposition 2 would still hold. The result of Proposition 2 is intuitively obvious and

---

[11]The mathematical formulation of the appropriate density function presented in this figure is from Johnson *et al.* The density function is a particular type of mean-preserving spread. It is used for a similar purpose later in this paper to derive a proposition with regard to uncertainty of enforcement of an ex ante regulation.

is not as insightful as that for Proposition 1. ***It should be made clear that neither of the above two propositions nor those that follow depend on the firm exhibiting risk-averse behavior. The firm is assumed to be risk neutral.***

Regarding assumptions concerning the behavior of the marginal cost function C'(x), we previously developed a third proposition regarding whether the firm would over- or under-protect against the environmental harm:

Proposition 3. Given a socially optimal level of precaution greater than zero, and a sufficiently large (small) marginal cost of precaution at the social optimum, x*, then the firm will employ too little (too much) precaution when facing strict liability.

Like Proposition 2 this proposition is intuitively obvious. The results stated in Propositions 1-3 are similar to those of Craswell and Calfee 1984 and 1986), who analyzed the behavior of a firm facing a negligence rule with uncertainty regarding the legal standard.

Earlier, we assumed that victims were perfectly compensated and the focus was on uncertainty regarding the determination of the proximate cause of the harm. Now we turn to the case where uncertainty regarding proximate cause is fixed but compensation is imperfect. The effect of imperfect compensation on the firm's precautionary decision can be stated as a fourth proposition:

Proposition 4. Assume that the firm's expectation of the costs of a harm is greater (less) than A(x), society's expectation regarding these costs. If so, then the firm will oversupply (undersupply) precaution when subject to strict liability.

This proposition is illustrated in Figure 4.

An example of a potential case in the environmental regulatory arena to which the above model and propositions might apply is that of the small-firm generator of hazardous wastes. Prior to the 1984 Amendments to the Resource Conservation and Recovery Act (ARCRA), small-firm generators of hazardous wastes faced little or no regulatory requirements in the handling of such wastes. These generators did however face public liability actions under the Comprehensive Environmental Response, Compensation and Liability Act (CERCLA) of 1982. But given the lack of record keeping regarding their disposal of hazardous wastes and therefore the difficulty of establishing causality, and problems with USEPA's Superfund Program there was a great deal of uncertainty regarding enforcement. Along with this great uncertainty were the relatively high marginal costs to these small firms of proper hazardous waste disposal and the reality that only public causes of action held any real chance of successful litigation. Given these factors, Propositions 1,3, and 4 would point to a high degree of under-precaution by these firms. Congress may have intuitively understood this when the small generator provisions in ARCRA were made considerably tighter.

B. A Model of the Firm Facing an *Ex Ante* Regulation

We now abandon strict liability for the moment to examine the case of the firm's behavior when facing an *ex ante* regulation. Because the deterministic model of the firm under these circumstances is well known, we will proceed immediately to the case of uncertain enforcement. Our task is greatly simplified by noting that the analysis of an uncertain ex ante regulation looks very much like that of an uncertainly enforced negligence rule, as described in Johnson *et al.* and Craswell and Calfee (1984 and 1986).

For the case of a fixed fine for violating the regulation the firm's objective function can be formulated as

$$\min_{X} \text{ Total Costs} = TC(x) = C(x) + F(y)P(x) \tag{6}$$

where C(x) is defined as the costs of precautionary measures to meet the regulation, F(y) is a fixed fine for not meeting the required level of precaution (y) ($F(y) > 0$ for $x < y$ and $F(y) = 0$ for $x \geq y$), and P(x) is the probability of being found in violation of the regulation. P(x) like R(x) in our earlier model of strict liability is given by

$$P(x) = \int_{x}^{\infty} g(x)dx \tag{7}$$

where g(x) is a continuous probability density with a support $[0,\infty)$. as before we will assume that $[C(X) + F(y)P(X)]$ is strictly convex. We will also asume that y is set equal to the social optimum. The first-order condition for the minimization of Equation 6 is

$$TC'(\tilde{x}) = C'(\tilde{x}) - F(y)g(\tilde{x}) = 0 \tag{8}$$

or

$$C'(\tilde{x}) = F(y)g(\tilde{x}).^{12} \tag{9}$$

An obvious question is how the distribution g(x) and the marginal cost function C'(x) will affect the choice of $\tilde{x}$ relative to the regulatory standard y. That is, can we tell whether uncertain enforcement of the regulatory standard punished by a fixed fine will induce optimal precaution, over-precaution, or under-precaution? We can determine this by evaluating

---

[12] The exposition here closely follows that of Craswell and Calfee and Johnson et al. for a firm facing a negligence rule with a fixed fine.

the sign of $TC'(\tilde{x})$ in Equation 8 when the firm's precaution is just equal to the regulatory standard. Because TC(x) is strictly convex and is minimized at $\tilde{x}$, TC'(y) < 0 when y < $\tilde{x}$ and TC'(y) > 0 for $\tilde{x}$ < y.

As in our discussion of the firm's behavior when facing a strict liability standard, let the exposition first turn to assumptions regarding g(x). To facilitate our comparative static analysis, we introduce a particular type of mean-preserving spread on g(x):

$$g_\gamma(x) = \begin{cases} \gamma g_\gamma[a(x-y) + y,], & \text{for } x \geq \frac{y(\gamma-1)}{\gamma} \\ 0 & \text{, otherwise} \end{cases} \tag{10}$$

where y is the expected value of x when x is distributed as g(x). It is easily seen that $g_1(x) = g(x)$. The density function $g_\gamma(x)$ is well-behaved for all values of $\gamma > 0$, and random variables distributed according to g and g_γ have the same mean. A decrease in γ causes the spread of g_γ to increase, and an increase in γ causes the probability mass to become concentrated at the mean similar to what occurs in Figure 2 for $q_a(x)$, a hypothetical distribution.

As uncertainty concerning enforcement of the ex ante regulation becomes greater (less) for the firm, i.e., as γ gets smaller (larger), g(y) becomes smaller (larger) causing TC'(y) in Equation 8 to become positive (negative). This implies that $\tilde{x}$ is less (greater) than y. This result can be made into a formal proposition.

Proposition 5. Assume that uncertainty regarding the enforcement of the regulatory standard y is distributed according to g(x) with g(x) > 0. If (in the sense of Equation 10) uncertainty is sufficiently large (small), then the firm subject to the standard will take too little (much) precaution.

The proof of this follows that for proving Proposition 1 given in Johnson *et al.* It can be shown that a sufficiently spread out (compact) version of g(x) exists such that TC'(y) in Equation 8 becomes positive (negative). Since $C'(x) > 0$, there exists a $\gamma > 0$ such that TC'(y) becomes positive and at least one value such that TC'(y) becomes negative.

Having the firm's expected value of precaution equal to the regulatory standard is not sufficient to insure that the level of precaution employed by the firm is in fact equal to that standard. This is true even for come risk-neutral firms. (Note the analogy between Propositions 1 and 5.)

This last proposition should convince the reader of the equivalency of an analysis of uncertain enforcement of a regulatory standard and that of uncertain enforcement of a negligence or strict liability rule. This will allow us to expedite the formulation of two more propositions regarding uncertain enforcement of a regulatory standard. The first of these regards biased perceptions of g(x):

Proposition 6. If the distribution g(x) is sufficiently biased in the sense of Figure 2 to the left (right) of y, and $q(y) > 0$, then the firm will take too little (little) precaution.

The proof of this proposition follows that given Johnson et al. for proving Proposition 2. While not proving a similar result, Craswell and Calfee show by means of simulation that the exact assumptions regarding the bias for either a shift in the mean or change to a non-symmetric distribution around the standard do not affect Proposition 6.

Next we will deal with the effect of marginal cost on the firm's choice of $\tilde{x}$ when there is uncertain enforcement of the regulatory standard.

Proposition 7. Given $g(y) > 0$, then for suitably large (small) marginal cost of precaution at the regulatory standard, y, the firm will employ too little (much) precaution ***vis-á-vis*** the standard.

(Note that Proposition 7 is equivalent to Proposition 3 presented earlier.)

By now the reader should be aware that Propositions 5 through 7 would hold if the firm faced a fine based on the level of precaution. The firm's objective function could be written as

$$\min_{x} TC(x) = C(x) + F(y-x)P(x) \tag{11}$$

where $F(y-x)$ is the fine function that meets the following conditions: $F(\cdot) \geq 0$ $(x \geq 0)$, equals zero when $x \geq y$, and $\partial F(\cdot)/\partial x < 0$.

We may summarize the results of this section in the following way: if any or all of the following conditions are true--there is great (small) uncertainty regarding enforcement of the regulation; there is a leftward (rightward) biased perception of the uncertainty surrounding enforcement of the regulation; or there are large (small) marginal costs at the regulatory standard, then Propositions 5 through 7 state that we would expect the firm to take too little precaution by comparison to the regulatory standard.

C. A Model of the Firm Facing Both *Ex Ante* Regulation and *Ex Post* Strict Liability

We now combine the separate analyses of uncertain enforcement of ex ante regulation and *ex post* liability in order to see under what conditions this simultaneous use of *ex ante* and *ex post* policies is efficient. We will first look at the case of uncertain enforcement of strict liability but certain enforcement of a regulatory standard.[13]

Introduction of the *ex ante* regulatory standard can be made by assuming that the firm's distribution around the social optimum is truncated. We assume that, given a regulatory standard, y, the firm will not take precautionary measures below y and that, therefore, the probability distribution around the social optimum is truncated. The result of this assumption is to move the probability mass that would be to the left of y without a regulatory standard to the right of y, i.e., there is a zero probability that the firm's choice of x will be below y. This will have the effect of raising the density function to the right of y and shifting the mean of the resulting probability distribution. The intuition behind this truncation assumption is that after the imposition of a regulatory standard, the firm perceives that the further its choice of x is to the right of y the less likely it will be found liable. This truncation assumption allows us to model the firm's behavior without a fine being associated with violation of the *ex ante* regulatory standard. Truncation alone is sufficient to affect the firm's behavior.

The new density function, $\underline{q}(x)$, that results from our truncation assumption has a conditional distribution

$$\underline{q}(x) = q(x \mid x \geq y). \tag{12}$$

The conditional probability $\underline{R}(x)$ that the firm will be found strictly liable if its level of precaution is x is

$$\underline{R}(x) = \frac{R(x)}{R(y)}. \tag{13}$$

---

[13]A certain regulatory standard can be thought of as having a point distribution regarding uncertainty of enforcement.

The firm's objective function becomes

$$\min_{x} TC(x) = C(x) + A(x)\underline{R}(x). \tag{14}$$

The first-order condition for this minimization is

$$TC'(\tilde{x}) = R(y)C'(\tilde{x}) + A'(\tilde{x})R(\tilde{x}) - A(\tilde{x})q(\tilde{x}), \tag{15}$$

were $\tilde{x}$ is now understood to mean $\tilde{x}(y)$.

To show the firm's response to changes in the regulatory standard requires total differentiation of Equation 15 and the evaluation of the sign of $d\tilde{x}/dy$. The result of the total differentiation is

$$\frac{d\tilde{x}}{dy} = \frac{q(y)C'(\tilde{x})}{R(y)C''(\tilde{x}) + A''(\tilde{x})R(\tilde{x}) - 2A''(\tilde{x})q(\tilde{x}) - A(\tilde{x})q'(\tilde{x})} \tag{16}$$

Since $[C(x) + A(x)\underline{R}(x)]$ is convex, the denominator of Equation 16 is positive. The numerator is also positive, which implies that $d\tilde{x}/dy$ is greater than zero. Therefore, increasing the regulatory standard, y, has the impact of increasing the level of precaution taken. The results of this analysis can be stated as a proposition (Johnson *et al.).*

Proposition 8. Given that the firm faces strict liability with uncertain enforcement, imposition of a certain *ex ante* regulatory standard will always induce more precaution by the firm. This will promote efficiency if the firm would under-protect without *ex ante* regulation and will exacerbate inefficiency if the firm would over-protect in the absence of *ex ante* regulation.

In principle the optimal level of the regulatory standard, y*, needed to induce the firm to choose x^* can be determined. However, the relationship between y* and x* can be shown by simply rewriting Equation 15 after substituting in Equation 2, as

$$C'(x^*)[R(y^*) - R(x^*)] - A(x^*)q(x^*) = 0. \tag{17}$$

Evaluation of this equation gives us another proposition.

Proposition 9. The optimal ex ante regulatory standard, y^*, given its certain enforcement and the uncertain enforcement of *ex post* liability, will be less than the socially optimal level of precaution, x^*, provided $q(x^*) > 0$. $y^* = x^*$ iff $q(x^*) = 0$.

The situation of a hazardous waste landfill firm facing both ARCRA and CERCLA can now be evaluated in light of Propositions 8 and 9. Under CERCLA the firm faces potential pre- and post-closure liability with uncertain enforcement for the cleanup of leaks from the site. If the landfill site is currently not a Superfund site, then the uncertainty regarding pre- and post-closure liability may be great. This uncertainty, coupled with the fact that in all likelihood the firm will only face public causes of action and that as a result there will be less than perfect compensation of victims, will result in the firm's providing too little precaution. However, under ARCRA the firm faces nearly certain enforcement of ARCRA regulations regarding the construction of the landfill site, and the handling and recording of wastes received. These *ex ante* regulations should force the firm to chose a higher level of precaution than it would under CERCLA alone and thus to move toward the social optimum. It is still an open question as to whether the regulations are sufficient to shift the firm's choice of level of precaution to the social optimum.

The final embellishment to the present model is to include the possibility that there is uncertain enforcement of both the ex ante regulatory standard and ex post liability. In adding uncertainty regarding the enforcement of *ex ante* regulation, we must look at the possible states of the world that the firm faces.

1. The firm will be found in violation of the regulatory standard and therefore found liable for all environmental harms stemming from such a violation.

2. The firm will not be found in violation of the regulatory standard but will be found strictly liable for the environmental harms.

3. The firm will not be found in violation of the regulatory standard nor will it be held strictly liable for the environmental harms.

The first two of these outcomes indicate a conditional relationship between being found in violation of the regulatory standard and being held strictly liable for alleged environmental harms. We can think of the firm facing a sequential process-- it is first determined whether or not it has violated the regulatory standard; then it is determined if it is strictly liable for environmental harms. If the firm is not in violation of the regulatory standard, then in the second step of the process it is assumed that its level of precaution is at least equal to the standard. This allows the retention of $\underline{q}(x)$ as the appropriate density function.

Given a fixed fine for violation of the regulatory standard, the firm's objective function can now be written as

$$\min_{x} TC(x) = C(x) + F(y)P(x) + A(x)\{P(x) + [1-P(x)]\underline{R}(x)\}. \tag{18}$$

As before we will assume that the right-hand-side of Equation 18 is strictly convex. The first-order condition for Equation 18 is

$$\begin{aligned} TC'(\tilde{x}) = {} & R(y)\{C'(\tilde{x}) + A'(\tilde{x})P(\tilde{x}) - g(\tilde{x})[F(y) + A(\tilde{x})]\} \\ & + R(\tilde{x})\{A'(\tilde{x})[1 - P(\tilde{x})] + g(\tilde{x})A(\tilde{x})\} \\ & + q(\tilde{x})P(\tilde{x})A(\tilde{x}) = 0. \end{aligned} \tag{19}$$

As was the case above we need to totally differentiate Equation 19 and determine the sign of $d\tilde{x}/dy$. The results of total differentiation, upon rearrangement of the terms, is

$$\frac{d\tilde{x}}{dy} = \frac{q(y)\{C'(\tilde{x}) + A'(\tilde{x})P(\tilde{x}) - g(\tilde{x})[F(y) + A(\tilde{x})]\}}{D(\tilde{x},y)} \tag{20}$$

where

$$\begin{aligned} D(\tilde{x},y) = {} & R(y)\{C''(\tilde{x}) + A''(\tilde{x})P(\tilde{x}) - 2g(\tilde{x})A'(\tilde{x}) - g'(\tilde{x})[F(y) + A(\tilde{x})]\} \\ & + A''(\tilde{x})R(\tilde{x})[1-P(\tilde{x})] - A'(\tilde{x})g(\tilde{x})[1 - P(\tilde{x})] + 2A'(\tilde{x})R(x)g(x) \\ & - 2A(\tilde{x})g(\tilde{x})q(\tilde{x}) + [P(\tilde{x}) - 1][q(\tilde{x}) + A(\tilde{x})q'(\tilde{x})] + q'(\tilde{x})A(\tilde{x})R(\tilde{x}), \end{aligned} \tag{21}$$

the second derivative of the total cost function TC(x).

$D(\tilde{x},y)$ is positive because Equation 18 was assumed to be strictly convex. The numerator of Equation 19 is also positive if the uncertainty surrounding the enforcement of the regulatory standard is in some sense small, which implies that $d\tilde{x}/dy$ is greater than zero.[14] Thus in this case, just as was the case for the regulatory standard given certain enforcement, increasing the minimally acceptable regulatory standard has the effect of increasing the precaution taken. This outcome follows from the fact that the determination of violation comes prior to the judgement of liability. Note that if $d\tilde{x}/dy$ 0, this would be a *prima facie* for the abandonment of strict liability because y* would need to be set greater than or equal to the social optimum, x*.

The implication of the above findings is that if $\tilde{x}$ < x* (i.e., $\tilde{x}(0)$ < x*) but greater than y prior to the imposition of the *ex ante* regulation, then the introduction of regulation will promote efficiency. If, on the* contrary, $\tilde{x}(0)$ > x*, then ex ante regulation with small uncertainty will

---

[14] The definition of small uncertainty here is that which will result in too much precaution regarding the regulatory standard, i. e., $\tilde{x}$ > y. If this is the case the term F(y) equals zero.

exacerbate the inefficiency that exists when there is uncertain enforcement of a strict liability rule.

Proposition 10. The imposition of an *ex ante* regulatory standard, given the existence of a strict liability rule, will promote efficiency if the firm would under-protect in the absence of the standard and will exacerbate inefficiency if the firm would over-protect without the standard. This only holds if the choice of $\tilde{x}$ is greater than y implying that there is relatively little uncertainty regarding the enforcement of the regulatory standard.

Finally we can explore the relationship between the optimal level of the *ex ante* regulatory standard, y*, and the socially optimal level of precaution, x*. From Proposition 10 we know that y* = 0 if and only if $\tilde{x}(0) \geq x^*$. For $\tilde{x} < x^*$, $y^* > 0$ will promote efficiency. What level of y will make $\tilde{x} = x^*$? The answer can be found by substituting x* for $\tilde{x}$ in Equation 19 and solving for y*. However, While the actual computation of y* can be computed, the regulators must know the cost of precaution for the firm and the characteristics of the density functions q(x) and g(x). Stopping short of actually solving Equation 19 for y*, we can rewrite it, using Equation 2, as

$$\begin{aligned} TC'(\tilde{x}) = {} & [R(y^*) - R(x^*)][C'(x^*) - C'(x^*)P(x^*)] \\ & - R(y^*)g(x^*)A(x^*) + R(x^*)g(x^*)A(x^*) \\ & + q(x^*)P(x^*)A(x^*) = 0.\text{[15]} \end{aligned} \tag{22}$$

For the equality with zero to hold, the term [R(y*) - R(x*)] must be positive. This implies that the optimum level of the regulation is less than or equal to the optimal level of precaution, i.e., $y^* \leq x^*$.

---

[15] Note that the term F(y) is equal to zero at x^* and does not appear in Equation 22.

Proposition 11. The optimal level of the *ex ante* regulatory standard, y^*, given that strict liability exists, will be less than the socially optimal level of precaution, x^*, provided $q(x^*) > 0$ and $g(x^*) > 0$.

Note that unlike the case of a certain *ex ante* regulatory standard There will be no case where $y^* = x^*$, because there is uncertainty surrounding the enforcement of y^*. It should also be intuitive that with uncertainty surrounding the enforcement of y^* being small in the sense used in Proposition 5, the distance between y^* and x^* will be greater with uncertain enforcement than with certain enforcement.

The implication of the above finding is that where the optimal precaution calls for the joint use of ex ante regulation and ex post strict liability the regulatory standard should be set below the socially optimal level of precaution.

IV. Conclusions

There are two groups of conclusions that can be drawn from the above research. The first group of conclusions is regarding the use of either an ex ante regulatory standard or *ex post* strict liability rule by themselves when either has uncertain enforcement. The second group of conclusions concerns the complementary use of *ex ante* regulation and ex post strict liability.

The first conclusion regarding the sole use of either *ex ante* regulation or *ex post* strict liability when neither has certain enforcement is that if the uncertainty has an expectation of the social optimum, x^* (or $y^* = x^*$), neither policy will result in an efficient outcome from society's point of view. With regard to either policy the firm will take to little (much) precaution if:

1. according to Propositions 1 or 5, depending on the policy, if firm's uncertainty regarding enforcement is too large (small);

2. according to Propositions 2 or 6, depending on the policy, if the firm's subjective distribution regarding uncertainty of enforcement is sufficiently biased below (above) the social optimum; and

3. according to Propositions 3 or 7, depending on the policy, if the firm's marginal costs at the socially optimal level of enforcement is large (small).

Furthermore, in the case of *ex post* strict liability, according to Proposition 4 under-(over-)estimation of victim compensation will result in too little (much) precaution. Given the great deal of uncertainty surrounding the monitoring and enforcement of *ex ante* regulation and the uncertainty surrounding the enforcement of *ex post* strict liability with regard to environmental problems, it appears that the sole reliance on either one policy or the other is unwise.

We can also derive important conclusions regarding the complimentary use of both *ex ante* regulation and *ex post* strict liability. First, when the joint use of these two policies is desirable then the optimal regulatory standard, y*, should be less than the social optimum. Also, if there is small uncertainty in the sense of Proposition 5 surrounding the use of an *ex ante* regulatory standard, the optimal level of y* should be set lower than that if there was certain enforcement. Finally, the *ex ante* regulatory standard should be used by itself if there is certain enforcement and the firm's density function, q(x), regarding the enforcement of the strict liability rule equals zero; or a movement of the *ex ante* regulatory standard toward x* does not result in a higher level of precaution on the part of the firm. This may point to the correctness of regulating hazardous wastes with both ARCRA and CERCLA.

Finally, the validity of the above modeling effort can only be proved regarding the real world by empirical work. The most suitable method for empirical analysis may be simulation as employed by Craswell and Calfee (1986).

Bibliography

Cooter, Robert D. and Thomas S. Ulen, "An Economic Case for Comparative Negligence," *New York University Law Review,* 1986.

Cooter, Robert D. and Thomas S. Ulen, *Law and Economics,* Glenview, IL: Scott, Foresman and Co, 1987.

Craswell, R. and John E. Calfee, "Some Effects of Uncertainty on Complience with Legal Standards," *Virgina Law Review,* 1984, 70.

Craswell, R. and John E. Calfee, "Deterrence and Uncertain Legal Standards," *Journal of Law, Economics, and Organization,* Fall 1986, 2(2).

Harvard Law Review, "Developments in the Law-Toxic Waste Litigation," May 1986, *99.*

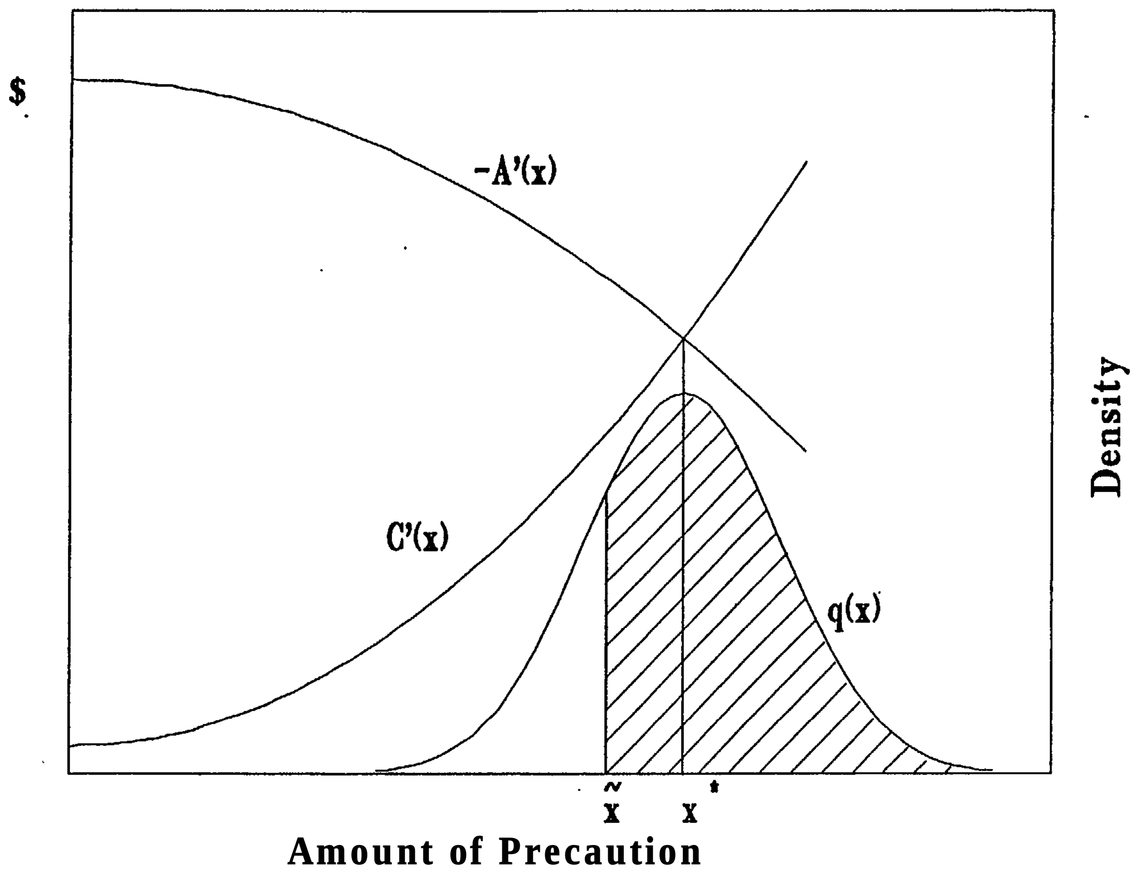
Johnson, Gary V., Charles D. Kolstad, and Thomas S. Ulen, "*Ex Ante* Regulation and *Ex Post* Liability: Substitutes or Complements?" (unpublished ms. 1987).

Johnson, Gary V. and Thomas S. Ulen, "Designing Public Policy Toward Hazardous Wastes: The Role of Administrative Regulations and Legal Liability Rules," *American Journal of Agricultural Economics,* 1986, 68.

Russell, Clifford S., Winston Harrington, and William J. Vaughan, *Enforcing Pollution Control Laws* (Washington, D.C.: Resources for the Future, 1986).

LEGENDS

Figure 1.
The Social Problem with Evidentuary Uncertainty for the Injurer



$
-A'(x)
C'(x)
q(x)
Density
x̃
x*
Amount of Precaution

Figure 2.
Two Levels of Uncertainty Around x*

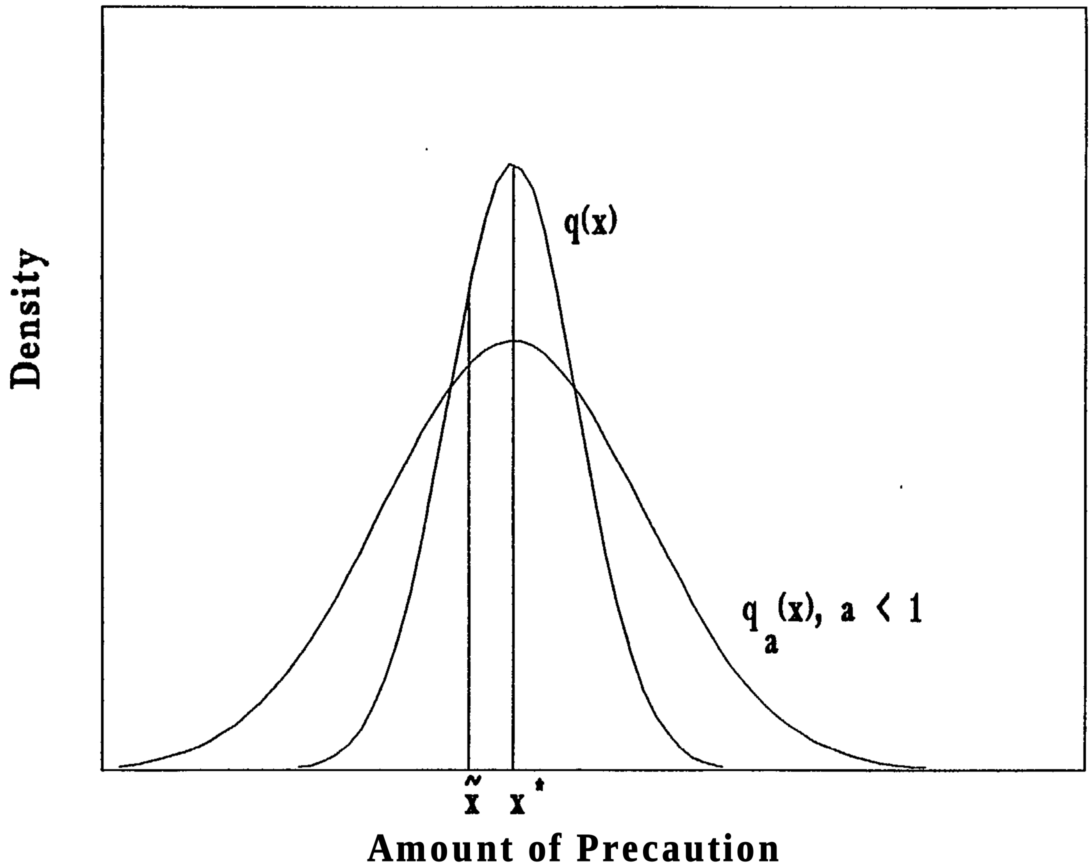

Density
q(x)
q_a(x), a < 1
x̃ x*
Amount of Precaution

Figure 3.
Biased Uncertainty

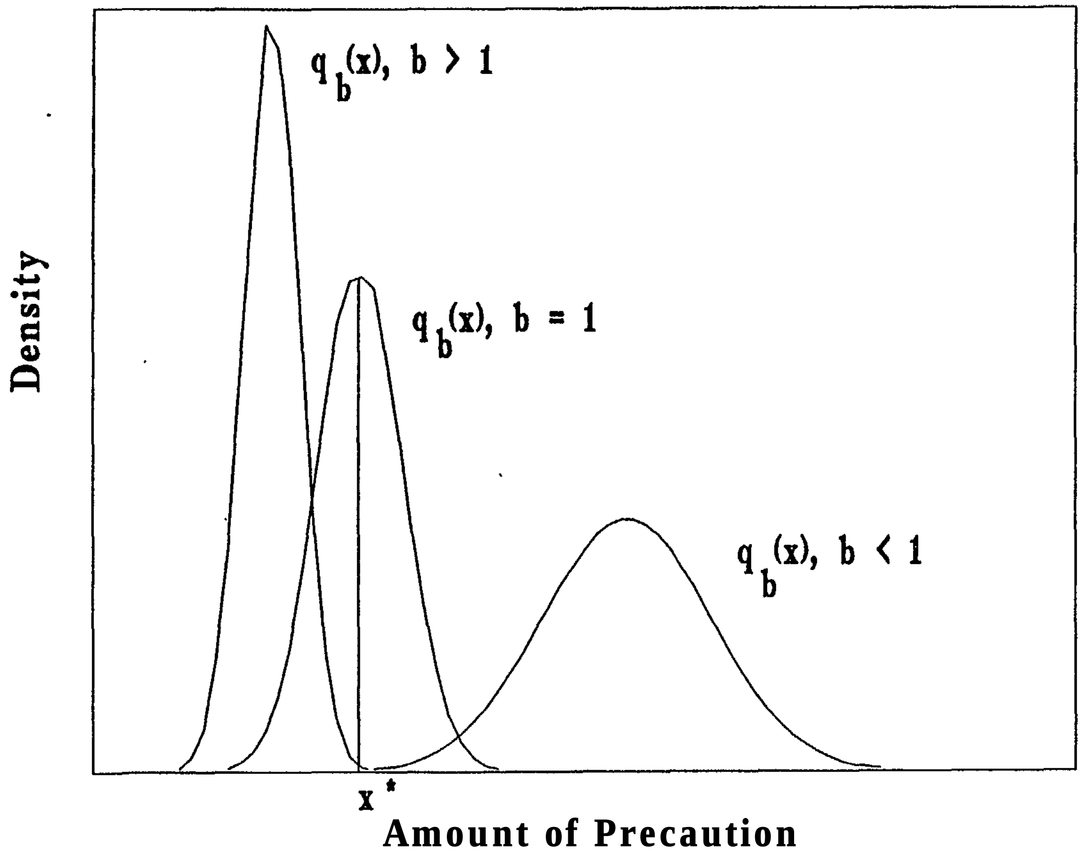
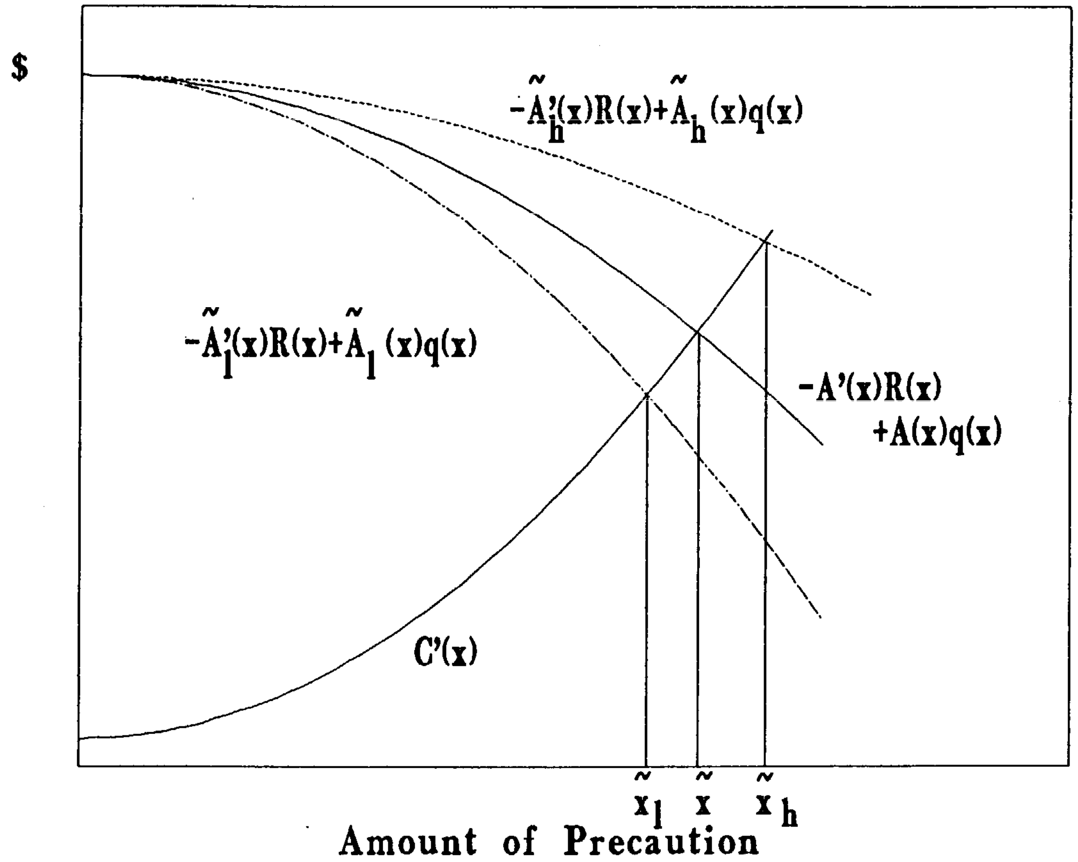

$q_b(x)$, $b > 1$
$q_b(x)$, $b = 1$
$q_b(x)$, $b < 1$
Density
x^*
Amount of Precaution

Figure 4.
The Injurer's Decision With Imperfect Compensation



$
$-\tilde{A}_h'(x)R(x)+\tilde{A}_h(x)q(x)$
$-\tilde{A}_l'(x)R(x)+\tilde{A}_l(x)q(x)$
$-A'(x)R(x)$
$+A(x)q(x)$
$C'(x)$
$\tilde{x}_l$
$\tilde{x}$
$\tilde{x}_h$
Amount of Precaution

MARKETS FOR POLLUTION CONTROL WHEN FIRMS ARE NONCOMPLIANT

Arun S. Malik

Bureau of Business and Economic Research
University of Maryland
College Park, Maryland 20742

July 8, 1987

1. Introduction

Over the past few years considerable support has developed for establishing transferable discharge permit (TDP) markets for controlling pollution. This interest in TDP markets is the result of a growing disenchantment with existing policies that rely on direct controls in the form of technology-based emissions standards. In addition to criticizing existing policies on the grounds of allocative inefficiency, many analysts have emphasized the problems of enforcing them in the face of widespread noncompliance.[1] Little attention has been given, however, to the consequences of noncompliance for the alternative policy instruments prescribed such as TDP markets.

The problems encountered in enforcing the current policy have been due largely to the technical difficulties in continuously monitoring pollutant emissions, and the absence of well-developed mechanisms for assessing penalties for noncompliance. As long as these deficiencies persist, noncompliance and the consequent need for enforcement will be an intrinsic part of any effective pollution control policy.

This paper examines the consequences of noncompliance for a TDP market. It does not attempt to characterize the optimal policy for enforcing such a market. Rather, it examines the effects of noncompliance given fairly general assumptions about the characteristics of the enforcement policy in place. This is accomplished by formally incorporating noncompliant behavior in a standard market model.

The first part of the paper examines questions such as whether TDP markets retain their efficiency properties in the presence of noncompliance, and what the effects of noncompliance are on the equilibrium permit price. The second part of the paper compares the consequences of noncompliance for a TDP market with those for an effluent tax policy, and discusses the problems that are likely to be encountered when implementing a TDP market given noncompliant behavior.

The existing literature on noncompliance and the enforcement of environmental policies is quite small. The first theoretical work is due to Harford (1978), who examines the behavior of a noncompliant, risk-neutral firm under two different policy instruments -- standards and taxes. A similar analysis is presented by Storey and McCabe (1980) for the case of a risk-averse firm. The only formal treatment of noncompliance in a TDP market is contained in a paper by Beavis and Walker (1983). They develop a simple model of a TDP market consisting of risk-neutral firms with stochastic pollutant discharge. The primary aim of their paper is to demonstrate that the frequency of monitoring does influence pollutant discharge when the latter is stochastic, contrary to what is allegedly found by Harford under the assumption that discharge is deterministic. As is shown below, Beavis and Walker incorrectly apply Harford's effluent tax model to a TDP market. Pollutant discharge is, in fact, a function of the frequency of monitoring whether discharge is stochastic or not.

A simplifying assumption made in this paper is that a firm's location is irrelevant: a unit of pollutant discharged by one firm is identical to that of any other firm insofar as its effect on ambient quality. Achieving a total discharge goal is then equivalent to maintaining a prescribed level of ambient quality, the usual context in which TDP markets are discussed. This assumption merely allows for an economy of notation; relaxing it does not materially alter the results presented.

2. The Model

The TDP market consists of n firms and a central authority responsible for issuing permits and ensuring that firms do not discharge more than the permitted amount. A total of s_T permits are issued by the central authority. These are allocated among the n firms via an auction or some other suitable mechanism. The authority may, for instance, initially distribute permits free, based on some equity criterion, and then allow firms to trade them. The firms are assumed to be price takers in the permit market and in their output markets.

The ith firm's profits from engaging in the pollutant generating activity, excluding permit and fine payments, are given by a smooth, strictly concave function $B_i(w_i)$ of pollutant discharge.[2] As is plausible, profits first increase in w_i, reaching a maximum at w_i^o, and then decline, as it becomes necessary for the firm to devote resources to generating pollution to further raise its discharge level. Clearly, the firm would not operate in the latter region.

The n firms face a common permit price r. Permit holdings are denoted by sit and the number of permits a firm initially receives free, if any, by s_i^o. If a firm is compliant, $w_i \leq$ Si and the magnitude of its violation, v_i, is zero. If a firm is noncompliant, $w_i > s_i$, and the magnitude of its violation is given by $v_i = w_i - s_i$.

Audit Probability and Penalty for Violations

The exact amount of pollutant discharged by a firm can be determined by the central authority only by conducting a compliance audit. It is assumed that these audits are unannounced and that firms cannot vary their permit holdings or pollutant discharge levels during an audit.

The central authority's decision on whether or not to audit a firm may depend on a variety of factors. These include the number of permits held by the firm (which is presumably public knowledge) and an estimate of the firm's pollutant discharge. The authority's ability to derive such an estimate, and its accuracy, will depend on the nature of the pollutant, its effects on environmental quality, and the characteristics of the firm's production process. For instance, particulate emissions from a smokestack can be estimated quite easily by remote monitoring. However, estimating discharges of water pollutants is likely to be difficult without conducting an on-site audit.

From the firm's perspective, audits are uncertain. This may be due to intentional randomness in the authority's audit policy or to uncertainty on the firm's part regarding the authority's estimate of pollutant discharges. Thus, a firm's subjective probability of being audited is given, quite generally, by $p_i(w_i,s_i;\theta)$, where e is a vector of audit policy parameters set by the central authority. Since the enforcement policy is taken as given, this parameter vector generally will be omitted. Note that the subscript on $p_i(.)$ implies that the audit probability functions are allowed to vary across firms. The presence of w_i in the probability function reflects the assumption that the authority's estimate of pollutant discharges is a function of actual discharges.[3] Assuming the estimate is an increasing function of actual discharges, we would expect the audit probability to be non-decreasing in w_i. The relationship between the level of permit holdings and the audit probability is more difficult to specify. It can be argued that the probability is increasing in s_i over some intervals and decreasing over others. I allow for both these possibilities.

It should be noted that the specification of the audit probability function includes the special cases of a constant audit probability, that is, one independent of w_i and s_i, and an audit probability dependent only on the firm's violation size, $v_i = w_i - s_i$.

If audited and found in noncompliance, the firm incurs a certain penalty, the magnitude of which is given by a smooth, increasing function of the firm's violation size $F_i(v_i;\delta)$, where δ is a vector of penalty policy parameters set by the enforcement agency. As with the audit parameters, the penalty parameters generally will be omitted below.

Firm's Decision Problem

Given the audit uncertainty, each firm seeks to maximize its expected utility of profits, $E[U_i(\pi_i)]$,

$$(1) \qquad \max_{w,\,s} \; [(1-p_i)U_i(\pi_i^0) + p_iU_i(\pi_i^1)]$$

where

$$\pi_i^0 = B_i(w_i) - r(s_i - s_i^0)$$

$$\pi_i^1 = \pi_i^0 - F_i(v_i)$$

and $p_i = p_i(w_i, s_i)$.

3. Optimal Pollutant Discharges and Permit Holdings

Pollutant Discharge

Omitting the firm subscript, the first-order conditions for an interior solution to (1) are:

$$(2) \qquad (1-p)U'(\pi^0)B' + pU'(\pi^1)(B' - F') = \frac{\partial p}{\partial w}[U(\pi^0) - U(\pi^1)]$$

$$(3) \qquad -(1-p)U'(\pi^0)r + pU'(\pi^1)(-r + F') = \frac{\partial p}{\partial s}[U(\pi^0) - U(\pi^1)]$$

Adding (2) and (3) and rearranging terms yields

$$(4) \qquad B_i'(w_i) = r + \phi_i \left[\frac{U_i(\pi_i^0) - U_i(\pi_i^1)}{E[U_i'(\pi_i)]}\right]$$

where $\phi_i = \partial p_i/\partial w_i + \partial p_i/\partial s_i$. The key expression in this equation is the second term on the RHS. Since $[U(\pi^0)-U(\pi^1)]/E[U'(\pi)] > 0$, the sign of the second term depends on that of ϕ_i. As noted earlier, $\partial p_i/\partial w_i$ is likely to be non-negative, but $\partial p_i/\partial s_i$ may be positive or negative. Hence, at this level of generality, the sign of ϕ_i is ambiguous. Note, however, that if

$$(5) \qquad \phi_i = \partial p_i/\partial w_i + \partial p_i/\partial s_i = 0$$

over the relevant range, (4) reduces to

$$(6) \qquad B_i'(w_i) = r,$$

which is identical to the abatement decision rule employed by a firm that is, by assumption, perfectly compliant. (A compliant firm maximizes $B_i(w_i) - r(w_i-s^0)$.)

Thus, if $\phi_i = 0$, the quantity of pollutant discharged by a noncompliant firm for a given permit price. is equal to that discharged by an otherwise identical compliant firm: $w_i^*(r) = \bar{w}_i^*(r)$, where $\bar{w}^*$ denotes the pollutant discharge of the compliant firm. An obvious corollary to this result is that when $\phi_i = 0$ the quantity of pollutant discharged by a noncompliant firm does not directly depend on the enforcement policy, or on the firm's attitude towards risk,

Although the condition $\phi_i = 0$ is fairly stringent, there are two interesting cases when it does hold. The first is when, over the relevant range, the subjective audit probability is a constant, independent of the firm's decisions, then, $\partial p_i/\partial w_i = \partial p_i/\partial s_i = 0$. The second case is when, over the relevant range, the audit probability is a function of the firm's violation size, that is, $p_i(w_i,s_i) = P_i(w_i - s_i)$; then, $\partial p_i/\partial w_i = P_i'$ and $\partial p_i/\partial s_i = -P_i'$. For these two cases, the abatement decision rule employed by a noncompliant firm (6) is identical to that used by a compliant firm.

In general, however, the presence of the second term on the RHS of (4) drives a wedge between the permit price (r) and marginal profits (B_i'). The sign of this term depends on the characteristics of the audit probability function. Its magnitude is determined, in part, by the firm's attitude towards risk, The strict concavity of the profit function implies that if the second term is negative, the quantity of pollutant discharged by a noncompliant firm is larger than that of an otherwise identical compliant firm $(w_i^* > \overline{w}_i^*)$. Conversely, if the second term is positive, the quantity of pollutant discharged is smaller $(w_i^* < \overline{w}_i^*)$.

Permit Demand and Equilibrium Permit Price

The above results present somewhat of an enigma. They indicate that, for a given permit price, the quantity of pollutant discharged by non-compliant firms my be equal to or even smaller than the quantity discharged by otherwise identical compliant firms. Yet, by definition, aggregate discharge must be higher in a TDP market in which firms are noncompliant. Upon reflection, it is apparent that for these two observations to be reconciled, the equilibrium permit price in a noncompliant TDP market must differ from that in a compliant market.

Examining the first-order condition for s_i (3), it is clear that a noncompliant firm's demand for permits, $s_i^*(r)$, depends on its attitude towards risk and on the characteristics of the enforcement policy it faces. Hence, the equilibrium permit price will also depend on these factors.

As is true for pollutant discharge, the relationship between a noncompliant firm's permit demand and that of an otherwise identical compliant firm depends on the characteristics of the firm's subjective audit probability function. As established above, if ϕ_i is non-negative, $w_i^* \le \bar{w}_i^*$. Since $w_i^* > s_i^*$ for a noncompliant firm, it follows that $s_i^* < \bar{w}_i^*$ when $\phi_i \ge 0$, that is, the permit demand of a noncompliant firm is lower than that of an otherwise identical compliant firm. (A compliant firm's permit demand is identical to its pollutant discharge level.) However, if ϕ_i is negative, $w_i^* > \bar{w}_i^*$. Although it is still true that $w_i^* > s_i^*$ for a noncompliant firm, this inequality no longer implies an unambiguous relationship between s_i^* and $\bar{w}_i^*$. Thus, it is possible when $\phi_i < 0$ that the number of permits demanded by a noncompliant firm is higher than that demanded by an otherwise identical compliant firm over some range of permit prices (i.e., $w_i^* > s_i^* > \bar{w}_i^*$).

The absence of an unambiguous relationship between the permit demands of noncompliant and compliant firms implies that no determinate relationship can be established between the equilibrium permit price in a market with noncompliant firms (r^*) and the equilibrium price in a market with otherwise identical compliant firms ($\bar{r}^*$). In particular, one cannot rule out the perverse possibility that the equilibrium permit price in a noncompliant market is higher than that in a compliant market. However, for the special case where ϕ_i is non-negative for all noncompliant firms in the TDP market, $r^* < \bar{r}^*$. This is shown below.

The equilibrium permit price in a compliant market is implicitly given by the equation, $\sum_i \bar{w}^*(r) = s_T$, whereas the equilibrium price in a noncompliant market is given by $\sum_i s_i^*(r) = s_T$. When ϕ_i is non-negative, $\bar{w}_i^* > s_i^*$, hence

$$\sum_i \bar{w}_i^*(r^*) > \sum_i s_i^*(r^*) = s_T.$$

Since $\bar{w}_i^*(r)$ is decreasing in r (see (6) and recall $B_i'' < 0$), it follows that $r^* < \bar{r}^*$. Thus, when $\phi_i \leq 0$ for all noncompliant firms (a condition which is automatically satisfied when audit probabilities are constant or a function of violations size), noncompliance results in a lower equilibrium permit price.

The above analysis indicates that the principal effect of noncompliance in a TDP market is to alter the equilibrium permit price. An important implication of this result is that noncompliance on the part of even one firm in the market will affect the pollutant discharge levels of all the other firms through its impact on the equilibrium permit price.

Some of the results derived thus far are illustrated in Fig. 1 for a TDP market consisting of two risk-neutral firms with audit probabilities that, for ease of exposition, are assumed to be constant for all values of w_i and s_i. Each firm's marginal profit curve (B_i') is depicted along with its marginal expected penalty curve $(p_i F_i')$. Firm 1's pollutant discharge and permit holdings are measured from the left side of the box, and Firm 2's from the right side. The distance between the sides of the box represents the total number of permits issued, s_T.

If we ignore the possibility of noncompliance, the equilibrium permit price, $\bar{r}^*$, and the equilibrium allocation of permits, A_c, are determined by the intersection of the marginal profit curves. Since the firms are compliant, A_c also gives their pollutant discharge levels.

If we allow for noncompliance and consider the marginal expected penalty curves faced by each firm, it is clear that Firm 2 will be compliant in equilibrium since its marginal expected penalty evaluated at v = 0 is larger than $\bar{r}^*$, which is an upper bound on the equilibrium permit price in the presence of noncompliance. Firm 1, on the other hand, will choose to be noncompliant.

Trading in the market amounts to sliding the 'branch" formed by the marginal expected penalty curves parallel to the horizontal axis. At the equilibrium depicted, Firm 1's marginal expected penalty is equated to its marginal profit, which, in turn, is equated to Firm 2's marginal profit. The common value of these three quantities is the equilibrium permit price, r*. Clearly, $r^* < \bar{r}^*$ given the assumption that the audit probabilities are constant.

The equilibrium allocation of permits in the presence of noncompliance is given by $\dot{A}_n$. Since Firm 2 is compliant, its pollutant discharge is equal to its permit holdings. However, Firm 1's discharge is given by $0_1 w_1^*$, which exceeds its permit holdings by the amount v_1. Both firms discharge more than they do in the absence of noncompliance.

Desirability of Equating Marginal Profits

In principle, when firms are perfectly compliant, TDP markets minimize the cost of achieving a given total discharge level since they ensure that marginal profits with respect to pollutant discharge are equated across firms. As we have seen, when firms are noncompliant, marginal profits may not be equated across firms. Only when the audit probabilities satisfy the condition in (5) over the relevant range do firms equate their marginal profits to the common permit price r. To conclude from this result that TDP markets may not be efficient when firms are noncompliant, it is necessary to establish that equating marginal profits is desirable even when firms are noncompliant.

In the perfect compliance case, the desirability of equating marginal profits is demonstrated by maximizing a social net benefit function of the form $[\Sigma B_i(w_i) - D(\Sigma w_i)]$, where D(.) captures the damages from pollutant discharges. The central authority is assumed to be able to directly control each firm's pollutant discharge. The first-order conditions for this problem establish that discharge levels are chosen such that marginal profits are equated across firms and set equal to the marginal damage from pollution.

The relevant benchmark maximization problem in the noncompliance case is considerably more complicated. It is now inappropriate to assume that the central authority can directly control each firm's pollutant discharge. However, it is reasonable to assume that the authority can indirectly control discharges by issuing non-marketable permits to firms: in effect, setting a discharge standard for each firm, Further control of discharges is provided by the authority's choice of enforcement policy. The objective function of the problem changes to the extent that it must also include the costs of enforcement.[4]

Characterizing the solution to the benchmark problem described above requires an analysis of the optimal enforcement policy, which is outside of the scope of this paper. The approach adopted here is to examine, instead, a sub-problem of the complete welfare maximization problem: one in which the enforcement policy is taken as given, and only the allocation of non-marketable permits is variable. The first-order conditions for the solution to this problem are a subset of those for the complete problem.

Since the enforcement policy is taken as given, enforcement costs are a constant and, for our purposes, can be omitted from the objective function. Hence, the objective function is identical in appearance to that for the perfect compliance case:

$$\max_{s_i} \sum_{i=1}^{n} B_i(\hat{w}_i^*) - D\left(\sum_{i=1}^{n} \hat{w}_i^*\right), \tag{7}$$

where $\hat{w}_i^* = \hat{w}_i^*(s_i)$ denotes a firm's optimal pollutant discharge level given a fixed number of permits, s_i. The first-order conditions for an interior solution to this problem are:

$$(B_i' - D')\frac{\partial \hat{w}_i^*}{\partial s_i} = 0 \qquad i = 1,\ldots,n. \tag{8}$$

These differ from the corresponding conditions for the perfect compliance case by the presence of $\partial \hat{w}_i^*/\partial s_i$. They imply, nonetheless, that it is desirable to equate marginal profits even when firms are noncompliant. We can therefore conclude that when firms are noncompliant, TDP markets are efficient only if $\phi_i = 0$ over the relevant range for all firms.

4. An Alternative Interpretation -- Effluent Taxes

A straightforward reinterpretation of the TDP model developed above allows us to examine the consequences of noncompliance for another pollution control policy widely advocated by economists--effluent taxes. To convert (1) to a model of a noncompliant firm facing an effluent tax, set s_i^o equal to zero, and redefine Si to be the reported discharge on which taxes are paid and r to be the unit effluent tax. Noncompliance now takes the form of tax evasion, that is, a firm under-reports its discharges. If the tax evasion is not detected, the firm pays taxes only on its reported discharge s_i. However, if the firm is caught, it also pays a penalty of

$F_i(v_i)$ based on the discrepancy between actual and reported discharge $v_i = w_i - s_i$.

The apparent identity between the model of a noncompliant firm in a TDP market and that of a tax-evading firm must be interpreted with caution. In the effluent tax interpretation of (1), the effluent tax, r, is presumably exogenous and not a function of firm behavior.[5] In contrast, in the TDP market interpretation, r is the equilibrium permit price which, as we have seen, is influenced by noncompliance and is endogenous. However, if we consider the firm in a TDP market in a partial equilibrium setting, there is no substantive difference between the TDP market and effluent tax interpretations. Hence the results derived for the TDP model when r is fixed are those that would obtain for a firm facing an effluent tax. In particular, we can state that a tax-evading firm sets its marginal profits equal to the effluent tax only if its audit probability function satisfies the condition $\phi_i = 0$ over the relevant range. If this condition is met, noncompliance does not affect the quantity of pollutant discharged by a firm; it only implies lower tax revenues.

The above results are a generalization of those obtained by Harford (1978) and Storey and McCabe (1980) using simpler models of the tax-evading firm. Beavis and Walker (1983) incorrectly conclude from Harford's tax model that even in a TDP market a firm's pollutant discharge is unaffected by changes in the enforcement policy when discharge is deterministic (as is assumed here). Their error stems from overlooking the indirect link between a firm's discharge decision and the enforcement policy via the equilibrium permit price.

5. Implementing a TDP Market when Firms are Noncompliant

Detecting Violations

The results obtained above suggest that a system of transferable permits may not be as robust to noncompliance as an effluent tax. As we have seen, under certain conditions, the only effect of noncompliance on an effluent tax policy is to lower tax revenues; pollutant discharges are unaffected. This is not true for a TDP market given the same set of conditions. Implicit in this comparison of the two policy instruments, however, is the assumption that the ease with which violations can be detected is similar for an effluent tax and a TDP market. As is argued below, it is likely that, in practice, detecting violations would be easier with a TDP market.

Given the technical difficulties in obtaining accurate, continuous records of pollutant discharge, enforcement agencies generally must rely on intermittent estimates of a firm's pollutant discharge.[6] Typically, these estimates are based on data collected during on-site audits. Hence, firms usually are aware of when their discharges are being monitored. This does not necessarily pose a problem in the case of a TDP market since the amount a firm should discharge is known beforehand by the agency from the firm's permit holdings. As with an effluent standard, any observed discharge over the allowed amount signals a violation. If firms receive advance notice of audits, they may be able to buy additional permits for the duration of the visit, making it difficult for the agency to detect violations. But the scope for such strategies would be limited if audits are unannounced (as is assumed in the model) or permits can only be traded at certain times. David et al. (1980) have suggested quarterly auctions as a means of precluding such behavior. Thus, in a carefully designed TDP market, detecting violations should be no moze difficult than under the existing policy of effluent standards.

Consider now the problem of detecting a violation when an effluent tax is employed. With a tax, the amount of pollutant a firm should discharge is determined when the firm reports (or formally records) its discharge, which occurs after the pollutant is discharged. As such, it is possible for tax-evading firms to adopt a strategy of accurately reporting discharges only when they are being monitored and thereby avoid being caught in flagrante delicto. Of course, there are limits to the extent to which this strategy would work. For instance, it is unlikely that firms could get away reporting near zero discharges when not monitored and levels substantially larger than zero when monitored. Nonetheless, the feasibility of this type of strategy implies that detecting violations is likely to be easier with a TDP market than with an effluent tax, given the existing technology for measuring pollutant discharges.

Containing the Effects of Noncompliance

A problem that is likely to be troublesome when implementing a TDP market is containing the effects of noncompliance, For an effluent tax or an effluent standard, changes in the compliance status of a particular firm do not necessarily affect the behavior of other firms: the effects of noncompliance are essentially localized. If a noncompliant firm revises its subjective audit probability as a result of, say, being caught and fined, this will not, per se, affect the decisions of other firms. Those that are compliant presumably will continue to be compliant, and those that are not may or may not alter their behavior. Of courser if the event is perceived as part of a larger crack down on noncompliance, the other noncompliant firms may revise their audit probabilities and alter their behavior accordingly.

In contrast, in a TDP market, the effect of noncompliance is transmitted throughout the market as a result of its influence on the equilibrium permit price. Changes in a single noncompliant firm's audit probability will, in general, alter the firm's demand for permits and, consequently, the equilibrium permit price. For a new equilibrium to be achieved, it may well be necessary for all firms, both noncompliant and compliant, to trade permits and adjust their discharge levels. This is especially true in markets with small numbers of dischargers, which are likely to be the rule rather than the exception.

Thus, given noncompliance, it may be difficult for a TDP market to achieve an equilibrium or maintain one, and firms may incur significant adjustment costs in modifying their discharge levels in response to changing permit prices. The significance of this problem will depend on the degree to which permit prices are affected by perceived changes in enforcement policy. In practice, permit prices may be far more sticky than models such as the one developed here lead us to expect.

Designing Trading Mechanisms to Accommodate Noncompliance

Undoubtedly, the stickiness of permit prices will depend on the mechanism adopted for" effecting permit trades. It has been implicitly assumed above that the TDP market takes the form of a standard unregulated market where trades occur whenever two or more parties find them advantageous. However, a number of policy analysts have argued that to mitigate problems of market thinness, TDP markets should take the form of periodic auctions conducted by a central authority (e.g., Noll, 1982; David, et al., 1980). A question that naturally arises is how noncompliance would be handled in such a market. Would a special auction be schedule whenever one or more firms are found in noncompliance to allow them to satisfy their possibly revised permit demands, or would they be required to wait until the next auction? In the former case, the auction-based market would function much like a conventional unregulated one. However, in the latter case, firms would effectively face a system of effluent standards between auctions, with each firm's standard given by its existing permit holdings. Hence, any efficiency properties of the TDP market, would be undermined between auctions. The significance of this problem clearly will depend on the frequency with which auctions are held, and the extent to which firms modify their discharge levels in response to perceived changes in the enforcement policy. It is likely that the problem can be mitigated in a carefully designed auction-based market.

6. Conclusions

Starting with a model of a noncompliant firm, I have examined the consequences of noncompliance for a system of transferable discharge permits. The analysis reveals that when firms are noncompliant TDP markets retain their efficiency property only under some fairly stringent conditions. The principal effect of noncompliance on the market itself is to alter the equilibrium permit price. Although it is likely that the equilibrium permit price is lower given noncompliance, it is difficult to rule out the possibility of a higher equilibrium price

Via its effect on the equilibrium permit price, noncompliance on the part of any one firm, or group of firms, influences the behavior of all other firms in the market. This raises some important questions regarding the ability of a TDP market to achieve and remain in equilibrium when firms are noncompliant. It also implies that careful consideration must be given to noncompliance when designing the mechanism for effecting permit trades.

REFERENCES

Beavis, B. and Walker, M. (1983), "Random Wastes, Imperfect Monitoring, and Environmental Quality Standards," *Journal of Public Economics,* 21:377-387.

David, H., Eheart, W., Joeres, E. and David, E. (1980), "Marketable Permits for the Control of Phosphorous Pollution into Lake Michigan," Water Resources Research, 16:263-270.

Downing, P.B. and Kimball, J.H. (1982), ‘Enforcing Pollution Control Laws in the U.S., " Policy Studies Journal , 11:55-65.

Drayton, W. (1980), "Economic Law Enforcement," *Harvard Environmental Law Review,* 4:1-40.

Harford, J.D. (1978), "Firm Behavior under Imperfectly Enforceable Pollution Standards and Taxes," *Journal of Environmental Economics and Management,* 5:26-43.

Noll, R.G. (1982), "Implementing Marketable Emissions **Permits,"** *American Economic Review, Papers and Proceedings,* 72:120-124.

Russell, C.S., Harrington, W. and Vaughan, W.J. (1986), Enforcing Pollution Control Laws, Resources for the Future, Washington, D.C.

Storey, D.J. and McCabe, P.J. (1980), "The Criminal Waste Discharger," *Scottish Journal of Political Economy,* 27:30-40.